

Penerjemahan dokumen inggris-indonesia menggunakan mesin penerjemah statistik dengan word reordering dan phrae reordering

Hansel Tanuwijaya, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=123771&lokasi=lokal>

Abstrak

Mesin penerjemah merupakan alat penerjemah otomatis pada sebuah teks dari satu bahasa ke bahasa lainnya. Tujuan dari mesin penerjemah adalah dapat membuat orang-orang yang berasal dari berbagai budaya, yang memiliki bahasa yang berbeda, dapat berkomunikasi satu sama lain dengan mudah. Mesin penerjemah statistik adalah sebuah pendekatan mesin penerjemah dimana hasil terjemahan dihasilkan atas dasar model statistik yang parameter-parameternya diambil dari hasil analisis korpus teks bilingual (atau paralel). Penelitian di bidang mesin penerjemah statistik untuk Bahasa Inggris-Bahasa Indonesia belum terlalu mendapat perhatian. Kualitas hasil terjemahan Bahasa Inggris-Bahasa Indonesia tersebut masih jauh dari sempurna dan memiliki nilai akurasi yang rendah.

Diawali dari permasalahan ini, munculah sebuah ide untuk membuat aturan-aturan restrukturisasi teks pada Bahasa Inggris sesuai dengan struktur Bahasa Indonesia dengan tujuan untuk meningkatkan kualitas dan nilai akurasi hasil terjemahan mesin penerjemah statistik. Aturan restrukturisasi teks tersebut bisa berupa word reordering, phrase reordering, ataupun keduanya. Dalam penelitian ini penulis merancang 7 buah aturan word reordering, 7 buah aturan phrase reordering dan 2 buah aturan gabungan phrase reordering dan word reordering.

Penelitian dilakukan dengan menggunakan Stanford POS Tagger, Stanford Parser, dan MOSES. Stanford POS Tagger digunakan dalam tahap word reordering, Stanford Parser dalam tahap phrase reordering, dan MOSES dalam tahap penerjemahan. Hasil eksperimen menunjukkan peningkatan akurasi dan kualitas penerjemahan yang efektif diperoleh dengan word reordering. Word reordering dapat memberikan peningkatan nilai BLEU sebesar 1.3896% (dari 0.1871 menjadi 0.1897) dan nilai NIST sebesar 0.6218% (dari 5.3876 menjadi 5.4211). Pada korpus bible, rata-rata nilai peningkatan nilai BLEU yang diperoleh dengan restrukturisasi teks adalah 0.5871% dan untuk nilai NIST terjadi penurunan sebesar 0.0144%. Pada korpus novel, rata-rata nilai peningkatan nilai BLEU yang diperoleh dengan restrukturisasi teks adalah 0.8751% dan untuk nilai NIST terjadi peningkatan sebesar 0.3170%. Besarnya peningkatan dan penurunan yang terjadi pada penelitian ini cenderung kecil (masih di bawah 1%). Hal ini dikarenakan aturan penerjemahan Bahasa Inggris-Indonesia menggunakan aturan MD-DM yang melibatkan penukaran kata yang jaraknya dekat sudah tercakup dalam distortion model pada mesin penerjemah statistik berdasarkan frase.

<hr>

Machine translation is an automatic translation tool for a text from one language to another language. The goal of machine translation is to allow people with different cultures and languages to communicate with each other easily. Statistical machine translation is an approach to machine translation in which the results produced on the basis of statistical model that its parameters taken from the bilingual corpus (or parallel)

text analysis. The research on statistical machine translation from English to Indonesian has not been received much attention. The English - Indonesian translation quality is still far from perfect and has low accuracy.

Based on this issue, come out an idea to make some text restructuring rules on English according to Indonesian language structure, with the purpose of improvement the quality and accuracy of the statistical machine translation. Text restructuring rules can be word reordering or phrase reordering or both. In this research, the authors design 7 word reordering rules, 7 phrase reordering rules and 2 combined phrase reordering and word reordering rules.

This research uses Stanford POS Tagger, Stanford Parser, and MOSES. Stanford POS Tagger is used in word reordering process, Stanford parser used in phrase reordering process, and MOSES in translation process. The results from experiments show that the most effective improvement is word reordering. The improvement with word reordering in BLEU score is 1.3896% (from 0.1871 become 0.1897) and for NIST score is 0.6218% (from 5.3876 become 5.4211). On bible corpus, the average of all text restructuring rules score are increased 0.5871% (BLEU) and decreased 0.0144% (NIST). On novel corpus, the average of all text restructuring rules score are increased 0.8751% (BLEU) and increased 0.3170% (NIST). The amount of increase and decrease that occurred in this study is considered as a small occurrence (which is still under 1%). This is caused by the MD-DM rules that involve exchanging words that have small distances between their range which have already been accounted for by the distortion model in phrase based statistical machine translation.